Comparison of speech intelligibility between English, Polish, Arabic and Mandarin

7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Comparison of speech intelligibility between English, Polish, Arabic and Mandarin

Kivanc Kitapci
School of the Built Environment, Heriot-Watt University, Edinburgh, United Kingdom.

Laurent Galbrun
School of the Built Environment, Heriot-Watt University, Edinburgh, United Kingdom.

Summary

In modern urban environments, the intelligibility of speech can be affected by cultural, lingual, and social diversity. The present study focuses on comparing four languages (English, Polish, Arabic and Mandarin) to examine how acoustic and linguistic factors influence the intelligibility of speech in a multilingual setting. Speech transmission index (STI) measurements and listening test results (diagnostic rhyme tests, phonemically balanced word lists and phonemically balanced sentence tests) were compared under four room acoustic conditions (STI = 0.2, 0.4, 0.6 and 0.8). The results obtained suggest that there is a significant difference between the word intelligibility scores of languages. English is the most intelligible language under all acoustic conditions, and differences with other languages are particularly large when room acoustic conditions are poor (STI = 0.2 and 0.4). Results also indicate that Arabic and Polish are particularly sensitive to background noise, and that Mandarin is significantly more intelligible than Arabic and Polish at STI = 0.4. Furthermore, a comparison between the word intelligibility scores and the sentence intelligibility scores shows lower variations between languages for the latter, and points out that each language has a different STI threshold above which the context of sentences becomes more obvious (i.e., the sentence intelligibility scores become higher than the word intelligibility scores). Amongst the four languages examined, this STI threshold is lowest for Polish and highest for English.

PACS no. 43.55.+p, 43.71.+m

1. Introduction

In a modern and globalised world, the interaction between multilingual and multicultural people in public, commercial and social spaces is gaining importance, and communication is at the center of this interaction. In the current literature, there are multiple studies which are looking at communication between non-native speakers; however, only very few studies have been comparing objective and subjective differences in speech intelligibility for native speakers of varying languages. The aim of the study presented is to find out possible relations between speech intelligibility and multi-lingual communication, in terms of acoustics, linguistics, and socio-cultural factors. In order to investigate the multi-dimensional structure of the intelligibility of speech in multi-lingual spaces, the project is divided into two main phases. The first phase investigates the interaction of various room acoustic parameters with different languages, and the second phase will investigate the role of socio-cultural aspects on the intelligibility of speech. In this paper, the results of the first phase are presented and discussed. The combination of the results obtained from both phases will lead to design guidelines and spatial design solutions for the use of service and product providers in order to minimise communication problems between end users.

Houtgast and Steeneken [1] investigated the correlation between various languages and speech intelligibility. They carried out a study using 11 western languages (English, Finnish, French, German, Hungarian, Italian, Dutch, Maori, Polish, Swedish and Slovak) in 16 acoustic conditions. As a result, it was found out that the differences among intelligibility tests may be caused by several effects, and that two of these effects are talker specific effects and phoneme or language specific effects [1].

One of the most relevant studies on comparing the intelligibility of different languages was conducted by Kang [2]. The intelligibility of English and Mandarin were compared in two spaces (a seminar room and a corridor) for three different room acoustic conditions. It was found that for a relatively high STI (high signal-to-noise ratio), the word intelligibility of Mandarin was better than English, and for a low STI, the intelligibility of English was better. It was also stated that Mandarin is slightly better than English under reverberant conditions, and English is considerably better than Mandarin under noisy conditions.

After reviewing the literature on both room acoustics and sociolinguistics, it was found that the number of studies that investigated the relationship between languages and speech intelligibility is limited. The present study aims to bridge that gap by comparing speech intelligibility of four languages (English, Mandarin, Polish, and Arabic) under various room acoustic conditions.

2. Methodology

This section describes the methodology that was used for selecting the languages, preparing the word and sentence lists, the recording procedure, and the listening test procedure.

Objective speech intelligibility was obtained from measurements of the speech transmission index (STI), which is a function of room acoustic properties and is based on the Modulation Transfer Function (MTF) method [3]. Subjective speech intelligibility was obtained from conducting word and sentence listening tests (diagnostic rhyme tests (DRT), phonemically balanced word tests (PB), and phonemically balanced sentence tests) which are typically based on the proportion of words correctly understood in a word or sentence list. Comparison of these results allowed identifying the correlations between subjective and objective speech intelligibility scores.

The study was carried out using several sample groups, in which the native language of each sample group was the variable. Languages representative of a wide range of linguistic properties were selected from different language families such as the Indo-European (e.g. English, German, Polish, Spanish, and Farsi), Uralic (e.g. Turkish), Afro-Asiatic (e.g. Arabic), Sino-Tibetan (e.g. Chinese) and Altaic (e.g. Japanese) language families. The specific languages identified for the research are English, Mandarin, Arabic, and Polish.

It should be noted that the selected language group were chosen to represent a western multilingual environment. Another important criterion used in the selection of the languages was the variability in consonant-to-vowel ratio, as the speech intelligibility is affected by the loss of consonants. Therefore, it was hypothesized that the languages that have a high consonant-to-vowel ratio might be more sensitive to the room acoustic conditions in terms of speech intelligibility. Another linguistic factor considered was the tonal properties of the languages. To examine the effects of the tonal system of a language on the speech intelligibility, at least one tonal language had to be selected. The native speakers' population of each language also had to be taken into account. The research should in fact be representative of a wide range of people; therefore, the languages with higher native speaker populations were selected. The availability of native speakers for the selected languages was also considered, and the languages selected had to comply with high number of participants that could be found at Heriot-Watt University.

Based on the above mentioned criteria of consonant-to-vowel ratio, tonal properties, and native speaker population, four languages were selected. These were English (low consonant-to-vowel ratio, wide-spread usage around the world), Mandarin (complex toned system, high native speakers' population), Arabic (moderately high consonant-to-vowel ratio, high native speakers' population), and Polish (high consonant-to-vowel ratio).

To assess the objective speech intelligibility, the speech transmission index (STI) method was used. To assess the subjective speech intelligibility, diagnostic rhyme tests (DRT), phonemically balanced word lists (PB) and phonemically balanced sentence lists were used. The DRT is a listening test consisting of 192 words arranged in 96 pairs [4]. The words are common, monosyllabic words, and most of them have three sounds ordered in a consonant-vowel-consonant sequence. The word pairs differ only in their initial consonants. DRT and PB tests were used to examine word intelligibility, whilst phonemically balanced sentence tests were used for the analysis of sentence intelligibility. It should be noted that

PB word tests were only used for Polish because of the lack of DRT material in Polish; however, the results are still comparable [4]. The word lists [4] [5] [6] [7] were recorded in an anechoic chamber using professional native speakers for each language (three males and three females) [4]. Because of the variety of accents of English and Arabic, attention was given to the origin of the talkers of these languages; therefore the English talkers were selected from south-eastern England, and the Arabic talkers were selected from Syria.

The words and sentences were then randomized in order to use the same lists several times for various acoustic conditions [4]. Before the actual recordings, a practice list was recorded by each talker, to make them familiar with the process.

The listening tests were conducted in one of the acoustic chambers of the acoustic laboratory of Heriot-Watt University. The dimensions of the chamber were 6.8m (l) x 4.0m (w) x 3.0m (h). All of the surfaces were reflective materials, and the room had no windows. Three male and three female listeners were selected from native speakers of each language. The recorded material was presented through a loudspeaker. Listeners' had to identify the spoken words within the word lists provided, whilst the sentences heard had to be written down. The listening test was repeated for four different acoustic conditions (STI = 0.2, 0.4, 0.6 and 0.8), by changing the reverberation time and signal-to-noise ratio. The reverberation time was controlled by mounting sound absorber panels on the walls (0.7s – 3.1s variation at 500 Hz), and the signal-to-noise ratio was controlled by adding white noise to the speech signal (S/N = +5 dB for STI = 0.4 and S/N = -5 dB for STI = 0.2).

The objective evaluation of speech intelligibility was measured using the commercial Maximum Length Sequence System Analyzer (MLSSA) software. The data gathered from MLSSA calculations were compared to the subjective speech intelligibility scores to examine potential correlations between the acoustic properties of a room and the subjective speech intelligibility, as well as correlations with the linguistic properties of a language.

3. Word intelligibility test results

This section examines correlations between subjective overall speech intelligibility results obtained from the diagnostic rhyme tests (DRT) and the objective speech transmission index (STI) measured under four room acoustic conditions. The results obtained from the preliminary analysis of the DRT results are presented in Figure 1. The horizontal axis shows the STI results, and the vertical axis shows the word intelligibility test scores for all languages. As stated previously, phonemically balanced word list (PB) results were used instead of DRT results for the Polish language.

Figure 1 illustrates that there are differences between subjective speech intelligibility scores of English, Polish, Arabic and Mandarin. First of all, English is the most intelligible language under all acoustic conditions. For the STI = 0.2 condition (S/N = -5 dB, high reverberation time) the DRT score of English is 37% and for the STI = 0.8 condition (no artificial background noise, low reverberation time) the DRT score is above 90%. It is also observed that Mandarin is more intelligible than Arabic and Polish at the STI = 0.4 condition (S/N = +5 dB, high reverberation time), in which participants were first introduced to the artificial background noise. The word intelligibility score of Mandarin at the STI = 0.4 condition is 69%, which is approximately 25% higher than the word intelligibility scores of Arabic and Polish. It is also seen that Arabic and Polish are the languages most sensitive to the introduction of artificial background noise. For Arabic, the difference of word intelligibility scores between the STI = 0.4 condition and the STI = 0.6 condition is 40%, and for Polish it is 46%. It is also apparent that the difference between intelligibility scores of languages becomes more conspicuous under poor acoustic conditions (STI = 0.4 and STI = 0.2). It is observed that there is an approximate difference between language scores of 9% for the STI = 0.8 condition; however, this increases to much larger differences of 33% for STI = 0.4, and 29% for STI = 0.2. It is also seen that there is a correlation between the consonant-to-vowel ratios of languages and the subjective speech intelligibility results for the most challenging room acoustic condition (STI = 0.2; high reverberation time and a low signal-to-noise ratio).

The differences in word intelligibility scores between languages were statistically analysed by the Factorial Analysis of Variance (ANOVA) method. This showed that there is a significant difference between the results of word intelligibility scores of each language, for each

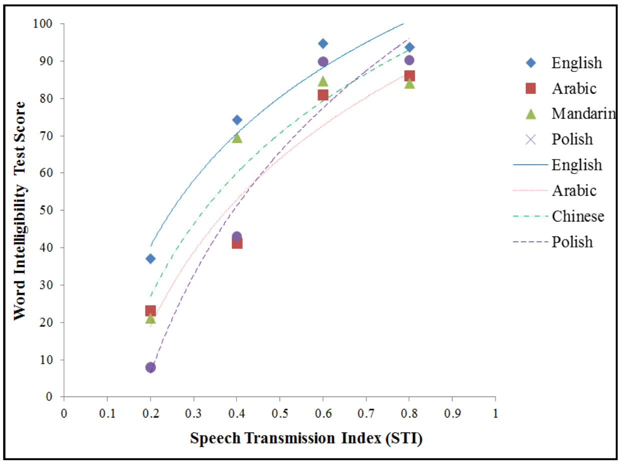

Figure 1 Comparison graph of word intelligibility scores and STI results for English, Polish, Arabic and Mandarin. Actual data markers and regression lines are shown in the figure.

room acoustic condition, which means that both the variation of languages (p = .000) and the variation of room acoustic conditions (p = .000) affect the intelligibility of speech.

The difference in results of word intelligibility tests between room acoustic conditions were statistically analysed for each language by using the one-way Analysis of Variance (ANOVA) method. The results of the one-way ANOVA were significant at the p = .000 level, indicating that the results obtained for each acoustic condition were significantly different.

4. Sentence intelligibility test results

In this section, phonemically balanced sentence test results for English, Polish, Arabic and Mandarin are presented and analysed. Sentence test scores were converted into percentages of correct scores, and the arithmetic average of all of the participants' results for each room acoustic condition was computed.

The comparison graph of the four languages' PB sentence test results is presented in Figure 2. Also, the comparison graphs of word and PB sentence scores for each language are presented in Figures 3, 4, 5, and 6. By looking at the trend lines created from individual PB sentences tests in Figure 2, it is seen that Arabic was significantly less intelligible compared to the other three languages. It should also be noted that at STI = 0.4 (high reverberation time, S/N = +5 dB) the variance of intelligibility is the largest. The difference between highest and lowest intelligible language at that point is approximately 40%. As stated in the interpretation of DRT results, Arabic has a high sensitivity to background noise, whereas Mandarin and English are less sensitive to background noise. At both ends of the trend lines, corresponding to STI = 0.8 and STI = 0.2, the intelligibility difference between languages is smaller than 10%. The difference between lowest and highest PB sentence test scores are larger at STI = 0.4 (~38%) and STI = 0.6 (~11%) compared to STI = 0.2 (~6%) and STI = 0.8 (~3%). Therefore, it can be stated that PB sentence tests are less accurate in identifying differences between languages when the acoustic condition is either challenging (STI = 0.2), or very good (STI = 0.8).

The results were statistically analysed by using the one-way Analysis of Variance (ANOVA) test for each participating language. Differences between acoustic conditions were statistically significant (p = .000), however differences between languages were not significant.

Further analysis of the sensitivity of PB sentence list was achieved by comparing the sentence and word intelligibility scores. The word intelligibility test vs. PB sentence score comparison graphs illustrate that there is a threshold where word and sentence intelligibility scores intercept (Figures 3, 4, 5, and 6). PB sentence tests tended to show higher intelligibility scores than the word tests above the threshold; however, below the threshold the word intelligibility scores were higher than the PB sentence test scores. The STI threshold value for the transition varied with language. For instance, for English the threshold was STI ≈ 0.6, for Polish it was STI ≈ 0.25, for Arabic it was STI ≈ 0.45 and for Mandarin it was STI ≈ 0.35.

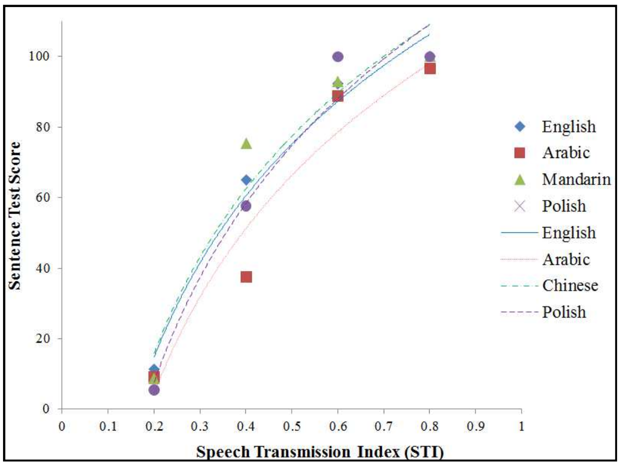

Figure 2 Comparison graph of sentence intelligibility scores (data markers and regression lines).

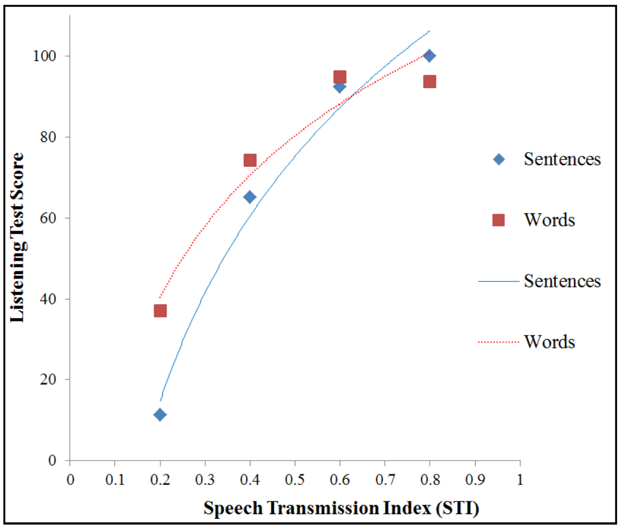

Figure 3 Comparison graph between sentence and word intelligibility scores for English (data markers and regression lines).

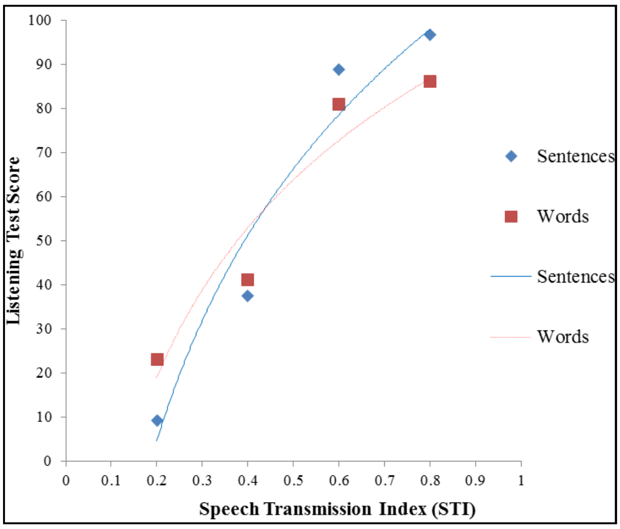

Figure 5 Comparison graph between sentence and word intelligibility scores for Arabic (data markers and regression lines).

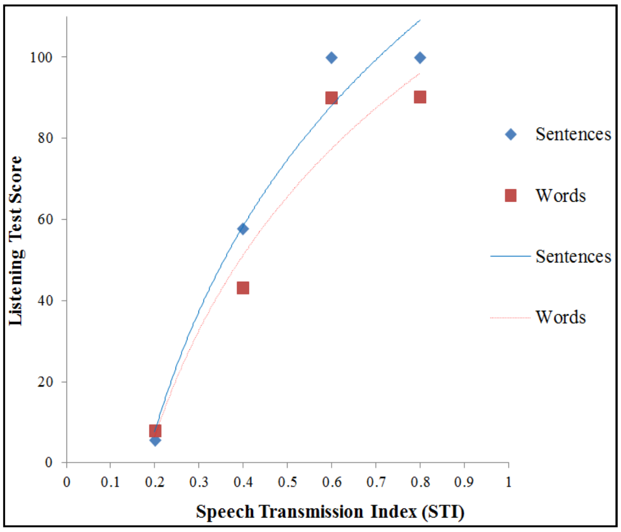

Figure 4 Comparison graph between sentence and word intelligibility scores for Polish (data markers and regression lines).

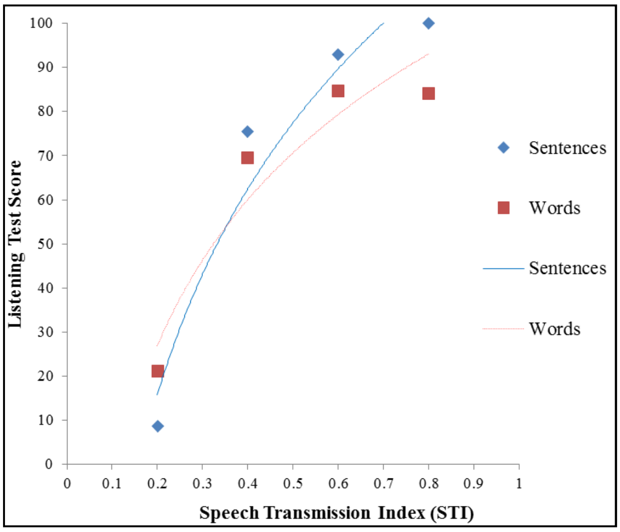

Figure 6 Comparison graph between sentence and word intelligibility scores for Mandarin (data markers and regression lines).

Apparently the difference between word and sentence intelligibility scores depends on the distance from the threshold value. The threshold can be interpreted as the STI level where context becomes intelligible enough. When the context becomes intelligible, even if not all the words can be understood, context can be transferred from the talker to the listener, and the sentences ultimately become 100% intelligible. Below the threshold, the boundary between syllables and words tends to disappear due to the high reverberation time and low signal-to-noise ratio. The lack of word and syllable boundaries decreases the overall intelligibility of speech [8]. Mandarin and Polish have a lower threshold compared to Arabic and English. Because of the varying thresholds observed for different languages, it can be suggested that there is no single optimum STI level for all of the languages.

5. Conclusions

The outcomes of the study revealed that there is a significant difference between the subjective word intelligibility scores of English, Polish, Arabic and Mandarin. Under the same acoustic conditions (background noise and S/N ratio), the word intelligibility scores of each language differ between each other, depending on the linguistic and consonantal properties of the languages. Also, a significant correlation was found between the consonant-to-vowel ratios and the word intelligibility scores of languages at the worst room acoustic condition (STI = 0.2). In contrast to word scores, sentences scores showed no statistically significant differences between languages. Additionally, the comparison between the word and the sentence intelligibility scores revealed that there is a language specific STI threshold, over which the context of speech becomes intelligible, therefore increasing the intelligibility of sentences. The data gathered in this phase of the research has been used to construct the second phase of the study, which investigates social and psychological effects on speech intelligibility.

Acknowledgement

The authors would like to thank Dr. Bernadette O'Rourke and Professor Graham H. Turner for the valuable help provided in the establishment and development of the research.

References

[1] T. Houtgast and H. J. M. Steeneken: A multi-language evaluation of the rasti-method for estimating speech intelligibility in auditoria. Acustica 54 (1984) 185-199

[2] J. Kang: Comparison of Speech Intelligibility between English and Chinese. J. Acoust. Soc. Am. 103 (1998) 1213-1216.

[3] T. Houtgast and H. J. M. Steeneken: The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 28 (1973) 66-73.

[4] ANSI/ASAS3.2: Method for measuring the intelligibility of speech over communication systems. American National Standards Institute (2009).

[5] B. Boudraa, M. Boudraa, and B. Guerin: Arabic diagnostic rhyme test using minimal pairs. Proceedings of Acoustics'08, Paris, France. (2008) 2329-2333.

[6] Z. Li, E. C. Tan, I. McLoughlin, and T. T. Teo: Proposal of standards for intelligibility tests of Chinese speech. IEE Proceedings Vision Image Signal Processing 147(3) (2000) 254-260 .

[7] E. Ozimek, D. Kutzner, P. Libiszewski, A. Warzybok, and J. Kocinski: The new Polish tests for speech intelligibility measurements. Proc. 2009 Signal Processing Algorithms, Architectures, Arrangements, and Applications (SPA), 163-168.

[8] A. Cutler, S. Butterfield: Word boundary cues in clear speech: A supplementary report. Speech Communication 147(10) (1991) 335-353.